DOCUMENT RESUME

ED 031 702 AL 002 047

By-Spolsky, Bernard

Reduced Redundancy as a Language Testing Tool.

Pub Date Sep 69

Note-17p.; Paper read to the Language Testing Section of the 2nd International Congress of Applied Linguistics, Cambridge, England, September 8-12, 1969.

EDRS Price MF-$0.25 HC-$0.95

Descriptors-Cloze Procedure, Language Proficiency, *Language Tests, *Redundancy, *Second Language Learning, *Testing, Test Validity, Transformation Generative Grammar

Identifiers-Clozentropy

Redundancy in language reduces the possiblity of error and permits communication where there is interference in the communicating channel. The relationship between creativity (the basic distinction between language-like behavior and knowing a language) and redundancy has been clearly established. Knowledge of rules is the key factor in creativity and is also the principal factor in the understanding of messages with reduced redundancy. In this paper the author discusses the three techniques that have been employed to test a subject's ability to function with a second language when noise is added or when portions of a test are masked. They are the cloze, clozentropy, and the noise tests. In the cloze test, portions of a written or oral test are blanked out and the subject is called on to provide the missing word or words. The clozentropy procedure measures a subject's performance in terms of a group norm. Thus, foreign students are scored according to the extent to which their responses agree with the normal responses of native speakers. The noise test originally consisted of a number of sentences to which noise had been added at signal to noise ratios of 1, 4, 7, 10, and 50 decibels. Recently, a multiple choice version has been prepared. The author feels that these practical studies have shown the value of language testing techniques based on redundancy. (DO)

ED031702

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Reduced redundancy as a language testing tool

Bernard Spolsky

The University of New Mexico

This paper has been prepared to be read to the Language Testing Section of the Second International Congress of Applied Linguistics, Cambridge, England, September 8-12, 1969.

AL002047

Language testing is a field that illustrates well the interdependence of applied and theoretical linguistics. In order to be able to develop valid tests of language proficiency, one needs to be able to find some way of characterizing the notion of knowing a language. What does it mean when we say that someone knows a language? This question turns out to be a central problem both of theoretical linguistics (the first task of which is to make explicit the complex set of systems that we call a language) and of applied linguistics as well. In that branch of applied linguistics which might be suitably called second language pedagogy, the question of knowing a language covers two areas of key importance: the specification of what is to be taught, and the testing of proficiency. Before we set out to teach someone a language we need to be able to say what it is that he should know at the end; and when we set out to test his proficiency, we are in fact involved in a definition of that knowledge.

A preliminary distinction must be made between knowing a language and language-like behavior. The key to this distinction lies in the notion of creativity. The creative aspect of language was for some time lost sight of in the behavioristic models that dominated linguistics in the first

half of the twentieth century. Chomsky (1964, 1966) has pointed out the two conflicting views of the essential nature of language held in the nineteenth century. On the one hand was the Humboldtian view: the essence of language is its form, a constant and unvarying factor underlying each new linguistic act. Contrasted with this was the view expressed by Whitney that language is "the sum of words and phrases by which any man expresses his thought". Saussure under Whitney's influence also regarded langue as an inventory of elements: it was perhaps for this reason that he relegated the sentence to parole. (Godel, 1966). The understatement of the creative aspect of language marked linguistics until Chomsky reformulated the Humboldtian position, tracing it back to Descartes. Descartes had noted the possibility of a machine which would give a specific number of responses to a specific number of cues, but he pointed out that one could not conceive of a machine that could reply appropriately to everything said to it, as a human being can. However imperfect a man is, he can arrange words together to express his thought; however perfect an animal, it cannot. The distinction is basic and not just connected with peripheral organs, for a parrot can utter words but cannot speak; a deaf-mute cannot produce words but can use language.

Consider the parrot. We can easily train him to produce a number of sets of sounds that seem like utterances. With the appropriate use of reinforcement, we can train him to produce each of these "utterances" on appropriate cues. We can train the bird to utter sounds that seem like "Please feed me" in order to receive food, or like "It is a pellet of food" when the food appears from the hopper. By the definitions of behavioral psychology these utterances could be classified as mand and tact respectively, essential elements of what Skinner (1957) has called verbal behavior. But I do not think that many of us would be prepared to call this behavior language. What is missing is the creative element. the parrot's repertoire of utterances remains limited and closed. We do not find it one day producing a new sentence. And the central fact in support of the creative aspect of language is that humans produce (and of course understand) many sentences that they have never heard before.

Creativity then is the basic distinction between language-like behavior and knowing a language. While precise specification may not be possible, for there is a continuum, the interpretation of each is relatively clear. Thus, language-like behavior refers to the parrot trained to speak, and equally well to the student who is able to

recite a number of sentences in a second language but not to modify them and use them in a free conversational situation. The example of the student learning a second language makes the continuum clear, for there is a stage at which he may be able to use his stock of sentences to answer a limited set of questions. This is still not the same as knowing a language, which involves the ability to produce an indefinite number of sentences in response to an indefinite number of stimuli.

One is said to know a second language when one's competence is like that of a native speaker. Performance need not however be identical, for it is accepted that someone knows a language when he speaks hesitantly, with many errors, or with a foreign accent. What confuses the distinction between language-like behavior and learning a second language is a third category, speaking a second language with the grammar of the first. It is normal for a person who knows one language and has developed language-like behavior in a second to be able to adjust his behavior in accordance with the grammar of the first. It is this that differentiates the human language learner from the parrot.

The creative aspect of language is one of the cornerstones of the argument for transformational grammar, for only such a grammar has available the "technical devices for expressing a system of recursive process" and only with

such devices can the creative aspect be formulated explicitly. (Chomsky, 1965) The only way to handle the fact that language has an infinite set of sentences and that it is used by people with finite time to learn it is to postulate a system of rules. Knowing a language is a matter of having mastered these rules, the ability to handle new sentences is evidence of knowing the rules that are needed to generate them.

Given this, the way to assess a subject's knowledge of the language is to test his ability to handle each of the rules of the language. But to do this requires a much more complete grammar of the language than is presently available. For all the great advances that have been made in grammatical description of English in the last decade or so, no one would be ready to claim that we have anything like a complete grammar available. For testing purposes, then, we have no way of determining the validity of the sample that we might choose. Our sample would necessarily be strongly biased towards those aspects that have so far been satisfactorily described. A second approach to the problem is however available, and may be defended both theoretically and practically. This is to make use of a second truth about language, that it is redundant.

Redundancy is a concept developed as part of the statistical theory of communication. (Shannon and Weaver, 1949)

In this theory, a message carries information to the extent that it effects a reduction in uncertainty by eliminating certain possibilities. The greater the reduction, the greater the information. Thus, the result of throwing a die (with six possibilities) is more information than the result of the toss of a coin (with only two possibilities). Or consider a linguistic example. Imagine that I ask someone to write down his first name. When I see him write the letter "p", my uncertainty has been reduced by a large amount, for he has excluded all names that begin with any other letter. When he adds the letter "a", uncertainty is further reduced as names like "Peter" and "Phillip" are ruled out. Adding the letter "u" makes it pretty easy to guess the final answer; "l" makes me almost positive, and his lifting the pen merely serves to confirm my guess. From this example, we see the way in which different parts of the message carry varying amounts of information. The letter "p" gave the most information, for it reduced the possibilities from the whole set of possible men's names to the set of names beginning with "p", a reduction, let us say, to 1/26 of the original. The letter "a" reduced the possibilities to an even smaller set, but by a smaller proportion. This is because in English only thirteen letters follow the letter "p". There just aren't any words in English that start with the letter "p" and have as their next letter

"b, c, d, g, j, k, m, p, q, v, w, x," or "z", so that the letter "a" in this second position reduces uncertainty to 1/13 rather than 1/26. As more letters are added, the amount of information conveyed by each letter becomes less, until certainty is reached. Now, the interesting thing here is the relation between the amount of information and our ability to guess. Our guessing, which gets easier as we go on (as the remaining elements contain less information) has depended on the knowledge that we have of the probabilities of occurrence of the various elements in the order they appear. It was our knowledge of the rules of English that permitted us to eliminate 50% of the possibilities for the second letter of the word. If English were not restricted in this way, the second letter would have conveyed as much information as the first. Probability relationships between the two letters, governed by the rules of English spelling and phonotactics, reduced the amount of information carried by the second letter. In a language without such restrictions, more information could be conveyed using fewer units. In natural languages, more units are used than are theoretically necessary; that is to say, natural languages are redundant.

Redundancy may seem wasteful of effort, but it is in fact of great use, for it reduces the possibility of error

and permits communication where there is interference in the communicating channel. When one considers all the interference that occurs when natural language is used for communication, it is clear that only a redundant system would work. Messages in normal language can be understood even though a good proportion of them is omitted or masked; or, in other words, every message contains many elements (defined statistically rather than linguistically) that can be omitted without leading to a break-down in communication.

But, if we give these distorted or incomplete messages to someone who doesn't know the language well, we find that there is a considerable difference. He needs the full normal redundancy, and at times even that is not enough. Note how when we speak to someone whose native language is different, we tend to speak more slowly, more clearly, with added gesture and frequent repetitions.

The relationship between creativity and redundancy has been clearly established. Knowledge of rules (the key factor in creativity) is also the principal factor in the understanding of messages with reduced redundancy. Miller and Isard (1963) have shown that the intelligibility of a sentence depends on its following syntactic and semantic rules. Sentences which break constraints (e.g., "A witness appraised the shocking company dragon") prove more difficult to

understand and repeat than those that do not, and ungrammatical sentences (e.g., "A diamond shocking the prevented dragon witness") prove even more difficult. This effect became even clearer when they studied the resistance of sentences to masking by added noise; grammatical sentences proved to be far more resistant than ungrammatical ones. Thus, they showed that the knowledge of the language providing the listener with help in handling sentences with reduced redundancy was a knowledge of rules, of the grammar of the language.

If we accept that knowledge of the language is a matter of knowledge of rules, what implications does this have for language testing? First, we must keep clear the many reasons for which language tests are designed; here we are concerned with proficiency tests, or what Lewis (1968) has called "summative assessment", and not with diagnostic tests. Further, we are concerned with a test that is independent of a specific set of language teaching materials and of the particular language analysis that lies behind it. In searching for a test of overall proficiency, we are trying to find some way to get beyond the limitation of testing a sample of surface features, or a sample from an incomplete description of the grammar, and are seeking to tap underlying linguistic competence. This can only be done with any degree of

certainty if we can be sure that we are presenting the subject with novel utterances, that is, utterances he has probably not heard before, or calling on him to produce novel utterances. The simplest way to do this is to set up an interview calling for normal language functioning. This method is however both difficult to score reliably and prohibitively expensive to administer. A more practical approach appears to be to make use of the principle of redundancy, and to test a subject's ability to function with a second language when noise is added or when portions of a test are masked.

There are three main techniques that have been employed to take advantage of this principle, the cloze, clozentropy, and the noise test. In the cloze test, portions of a written or oral test are blanked out and the subject is called on to provide the missing word or words. The masking is statistically controlled. Studies such as those by Taylor (1957), Carroll (1959), Holtzman (1965), and Weaver (1963) have shown this to be a reasonably good measure of language proficiency, but a major problem remained in the establishment of a scoring procedure. The question to which no satisfactory answer was found was what is to be counted as a correct response, the original deleted word, or any word that makes sense. Taylor reported correlation between the two methods

of scoring with native speakers, but Holtzmann reports no such correlation in the case of foreign students. Carroll raises the question that there might be an ability to handle the cloze unrelated to language ability, although it will need further research to determine whether this is not the factor of knowing how to do the test that plagues all testing. A recent study by Bowen (1969) has made use of the cloze technique in testing Ethopian students' relative control of Amharic and English. The value of the technique in this case was the possibility of preparing tests in the two languages of roughly equivalent difficulty without being forced to attempt to develop a notion of what is and what is not a difficult item in each of the two languages.

A new proposal by Darnell attempts to handle the difficulty of scoring. The clozentropy procedure, as it is called, measures a subject's performance in terms of a group norm. Thus, foreign students are scored according to the extent to which their responses agree with the normal responses of native speakers. The possibilities of the procedure are obvious. The tester can choose a sample population (say native speakers of English in the first year of a university engineering course) and claim to show how well a subject (say a foreign student applying for admission) would fit into this group. This more precise functional definition is likely to be of great value. Darnell reports very satisfactory reliability,

and high correlation of the instrument with TOEFL, the Test of English as a Foreign Language administered by the Educational Testing Service. Interestingly, the best correlation of the written clozentropy test appears to be with the listening comprehension section of the TOEFL test. The listening comprehension section of TOEFL is probably the part of the test that is least biased in favor of selected items, that is to say the part that is most likely to measure overall proficiency.

The third technique, the noise test, is one that I have worked on in the last few years. (Spolsky et al. 1968) In its first form, it consisted of a number of sentences to which noise had been added at signal to noise ratios of 1 db, 4 db, 7 db, 10 db, and 50 db. Mixing was done electronically, the signal being levelled by compression as necessary. This dictation test proved relatively easy to administer, not hard to score, and to correlate well with other tests of English proficiency. Results indicated that the test was successful as a device for screening foreign students entering an American University. The test sorted out those students who required no further English instruction and those whose ability was below that allowed for in the lowest level courses offered. Additional testing could then be concentrated upon those students for whom finely discriminated

levels had to be established. To simplify scoring, a multiple choice version has been prepared. The distractors used for each of the sentences were chosen according to errors made while the test was used in its dictation form. The administration of the test has thus been greatly simplified, and it has been possible to carry out item analysis.

These practical studies, then, have shown the value of language testing techniques based on the principle of reducing redundancy, and support Holtzman's suggestion that redundancy utilization might well turn out to be the integrative factor in language proficiency testing. Tests of this kind appear to be reliable, and relatively straightforward to prepare, administer, and score. Their validity has been shown by correlation with other tests and can be even more strongly supported by the theoretical arguments that I have been presenting.

Davies (1968) has presented a useful scheme for language test validities. He recognizes five types: face validity, based on a lay view of language; content validity, related to syllabus; construct validity, dependent on a theory of language learning; predictive validity, to be shown by later tests; and concurrent validity, shown by correlation with other tests. The face validity of tests using redundancy reduction can be established by pointing out that the

ability to understand speech masked by noise is clearly related to the ability to understand announcements in a noisy airport, or to understand a telephone conversation, something recognized by the layman as evidence of control of a foreign language. Content validity is clearly irrelevant for a test that claims to be independent of any syllabus, set of materials, or single language analysis. The tests have construct validity, not only in the sense allowed by Davies, in that they correspond to a theory of language learning, but even more in that, as I have shown, they accord to a theory of language and its related notion of knowledge of a language. Predictive validity raises more difficult problems, for it must be pointed out that language proficiency tests have generally not turned out to be good predictors of other behaviors. (Spolsky 1967, Upshur 1967) Finally, tests using reduced redundancy have been shown to have concurrent validity in that they correlate with a number of other types of measure of language proficiency.

Language tests based on these techniques may then be reasonably claimed to provide good measure of overall proficiency in a second language.

References

1. J. Donald Bowen. A tentative measure of the relative control of English and Amharic by eleventh-grade Ethiopian students. Workpapers in English as a second language. University of California, Los Angeles. 1969.

2. John B. Carroll, Aaron S. Carton, and Claudia P. Wilds. An investigation of "cloze" items in the measurement of achievement in foreign languages", Harvard University (mimeo, 1959).

3. Noam Chomsky. Current issues in linguistic theory in The Structure of language, edited by J. A. Fodor and J. J. Katz. New Jersey: Prentice-Hall, 1964.

4. _____. Aspects of the theory of syntax. Cambridge, Mass.: MIT Press, 1965.

5. _____. Cartesian linguistics. New York: Harper & Row, 1966.

6. Donald K. Darnell. The development of an English language proficiency test of foreign students, using a clozentrophy procedure. Final Report Project # 7-8-010, Grant OEG-8-8-070010-200(057), October 1968.

7. Alan Davies. Language Testing Symposium. London, Oxford University Press, 1968.

8. Robert Godel. F. de Saussure's theory of language, in Current Trends in Linguistics, Volume III, edited by Thomas A. Sebeok. The Hague: Mouton & Co., 1966.

9. Paul Holtzman. English language proficiency testing and the individual. Selected Conference Papers of the Association of Teachers of English as a Second Language. Los Altos, California. Language Research Associates Press, 1967. 76-84.

10. E. Glyn Lewis. International education assessment: English as a foreign language. Language Learning Special Issue No. 3, 1968, 125-45.

11. George A. Miller and Stephen Isard. Some perceptual consequences of linguistic rules. Journal of Verbal Learning and Verbal Behavior. 2 (1963) 217-28.

12. Claude E. Shannon and Warren Weaver. The mathematical theory of communication. Urbana: University of Illinois Press. 1949 (1963)

13. B. F. Skinner. Verbal behavior. New York: Appleton-Century-Crofts, 1957.

14. Bernard Spolsky. Do they know enough English? Selected Conference Papers of the Association of Teachers of English as a Second Language. Los Altos, California: Language Research Associates Press, 1967. 30-44.

15. _____, Bengt Sigurd, Masahito Sato, Edward Walker, and Catherine Arterburn. Preliminary Studies in the development of techniques for testing overall second language proficiency. Language Learning, Special Issue No. 3, 1968, 79-101

16. Wilson S. Taylor. Cloze procedures. Journalism Quarterly (1963) 415-33.

17. J. A. Upshur. English language tests and prediction of academic success. Selected Conference Papers of the Association of Teachers of English as a Second Language. Los Altos, California: Language Research Associates Press, 1967. 85-92.

18. Wendell W. Weaver and Albert J. Kingston. A factor analysis of the cloze procedure and other measures of reading and language ability. Journal of Communication. 13 (1963) 252-61.